Honors Thesis
A Study of Traffic Flow Using GPS Data

By:
Mark Lehman
lehman.155@osu.edu

Abstract

The purpose of this paper is to relate findings on the topic of highway traffic research. Specifically, GPS data were collected and examined to gauge traffic congestion on I-71. The location of the congestion as well as the cause of the congestion was examined. The minimum amount of data required to obtain similar results from a different section of roadway was quantified.

List of Figures

Introduction

Freeways are an important means of transportation. As freeways are used by an increasing number of people, they become congested. Congested freeways cause wasted fuel and frustration to people trying to traverse the congested areas. One method of studying congested freeways is using information from a vehicle traveling through the congestion. The information collected can be used to help alleviate future congestion and plan roadway improvements.

Probe Vehicle Data Analysis

The source of the traffic data used in this study is an otherwise normal vehicle traveling through traffic, called a probe vehicle, equipped with a DGPS (Differential Global Positioning System) [1] sensor and laptop. After the receiver processes the raw DGPS data, the probe vehicle's velocity and latitude/longitude coordinates are stored every second. The probe vehicle is driven along I-71 in Columbus several times per week during the peak usage hours of the roadway. Morning peak hours are taken to be 7-9 AM, while evening peak hours are 4:30-6:30 PM. The route driven is always the same north/south loop through Columbus, as circled in figure 1. The vehicle's drivers are instructed to drive in the middle lane of traffic, but to pass vehicles traveling significantly slower than the rest of the traffic, such as semitrailer trucks.

Analysis of probe vehicle data has been done by others prior to this study. A cost based analysis of using GPS equipped probe vehicles' data can be found at [3]. Time travel prediction using historical travel time values is presented in [4]. The main difference between this paper and others, including [4], is that the method of traffic analysis in this paper does not require any difficult to collect parameters about traffic flow and can be done with a relatively small amount of GPS probe vehicle data.

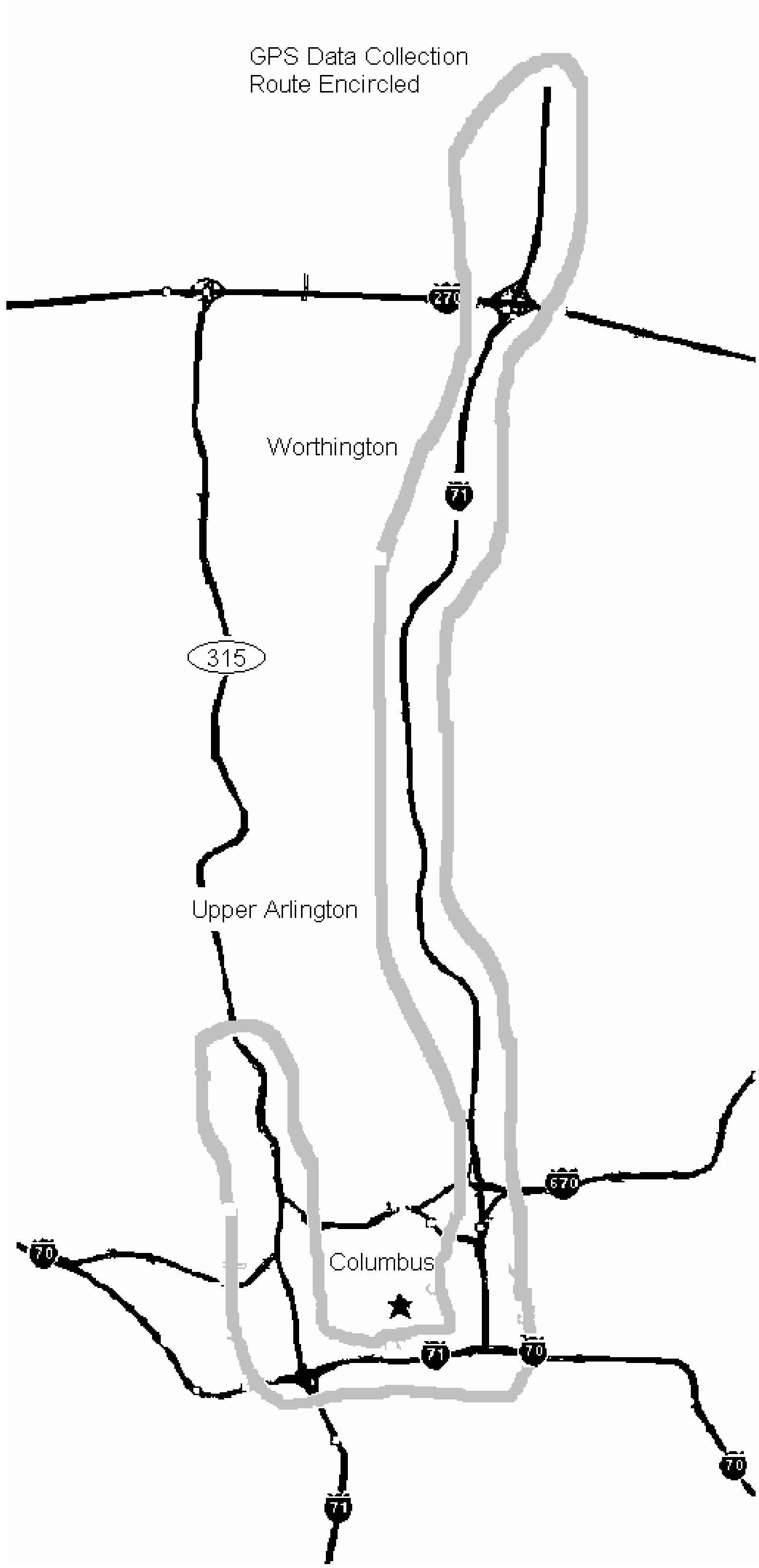

Figure 1: Data Collection Route through Columbus, Ohio. Adapted from [2]

Before the GPS data can be utilized, velocity measurement noise must be filtered out. When the van's communication with one or more GPS satellites is obstructed, usually when the van travels under a bridge, the velocity readings tend to be of poor quality. In order to detect velocity errors, the van's acceleration is approximated by taking the difference of successive velocity readings. If the magnitude of the acceleration is too large to be feasible, the data points are flagged as bad. Then, the bad points are replaced with a median of the adjacent samples. Several bad data points and the replacement median points are illustrated in the speed time series of figure 2.

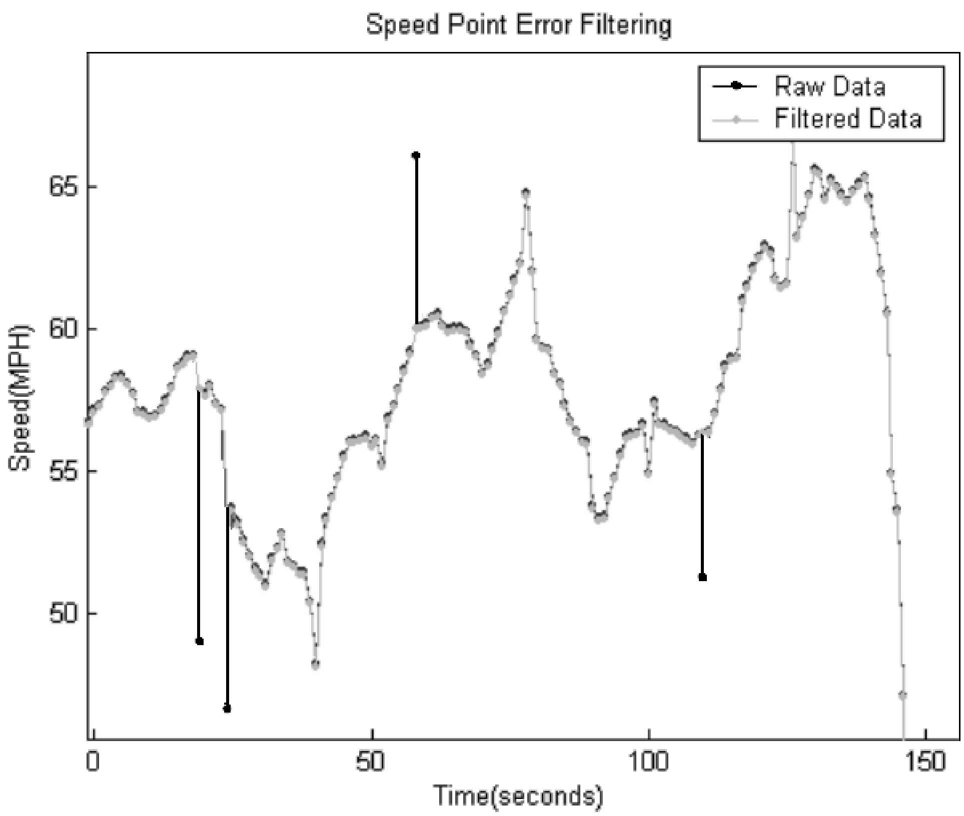

Figure 2

Once the data have been filtered, the latitude and longitude coordinates are converted to distance along the probe vehicle's route. Because the distance of each probe vehicle run is slightly different due to lane change maneuvers and other factors, direct comparison of the runs based on distance would be difficult if raw distance was used. Instead of converting latitude/longitude coordinates directly to distance, the coordinates are first snapped to a reference run. The latitude/longitude coordinates of the reference run are used to generate a distance along I-71. Each data point from a run is placed on top of the reference run, and the data points on the reference run are used to generate a distance for the data points on the rest of the runs. It should be noted that there is nothing special about the reference run; it is only used to make sure that all latitude/longitude coordinates correspond to the same distance along the run. After this normalization, the velocity readings from all the times the probe vehicle has been driven on the data collection route can be compared based on the location of the velocity readings.

Because traffic speeds on I-71 are typically above 50 MPH and probe vehicle drivers are instructed to pass vehicles traveling slower than the rest of the traffic, a state of congestion can be inferred from recorded speeds slower than 50 MPH. Comparing the velocity of many probe vehicle runs allows areas which regularly experience congestion to be located. When several probe vehicle runs travel through the same congested portion of roadway, the bottleneck that is causing the congestion can be identified. Knowing where congestion frequently occurs and the location of bottlenecks is important because it allows roadway planners to construct improvements which allow higher traffic flow.

Figure 3 shows probe vehicle runs passing through the main bottleneck on northbound I-71 in Columbus near 11th Avenue. Each of the curves on the plot is a different probe vehicle run. As the probe vehicle approaches the bottleneck, moving left to right on the plot, some of the runs experience congestion. The runs which do not experience congestion stay at a high speed (greater than 50 MPH) while the congested runs drop to significantly lower speeds, some even stopping briefly. At about mile 6.5, nearly all the congested runs begin to pick up speed. Dividing these runs into 3 groups, figure 4 shows all the probe vehicle runs which do not become congested at the 11th Avenue bottleneck. The important feature of figure 4 is that after mile 6, nearly all the data points are above 50 MPH. Figure 5 shows all the runs which do become congested when the bottleneck was active. The important features of figure 5 are that at some point the runs drop speed and enter congested traffic. Once the probe vehicle passes the 11th Ave. exit ramp, all the runs begin to pick up speed and eventually enter free flowing traffic. It is important to note that it does not matter where a run enters the congested traffic, they all pick up speed at the same location. This phenomenon indicates that a fixed feature on the roadway is causing the slow traffic while the length of the resulting queue varies. The traffic cannot instantaneously accelerate to free flowing, but instead picks up speed over a finite distance of roadway, an acceleration zone. In this acceleration zone, roughly mile 6.7 to 7, traffic exhibits features of both congested traffic and free flowing traffic. Slow and go waves, where drivers are forced to apply their brakes and then accelerate, are a prominent feature of congested traffic, but can still occur in the acceleration zones. After enough distance has passed, the runs which were formerly congested become indistinguishable from the uncongested runs. The congested traffic picking up speed and traveling at free flow speed indicates that the controlling bottleneck has been passed through. This particular bottleneck corresponds to a location on I-71 where the number of lanes drops from 4 to 3. Finally, a few of the runs do not pick up speed after passing the 11th Ave. bottleneck, shown in figure 6. The runs remain congested, meaning that a bottleneck farther downstream is restricting the flow of traffic more than the 11th Ave. bottleneck.

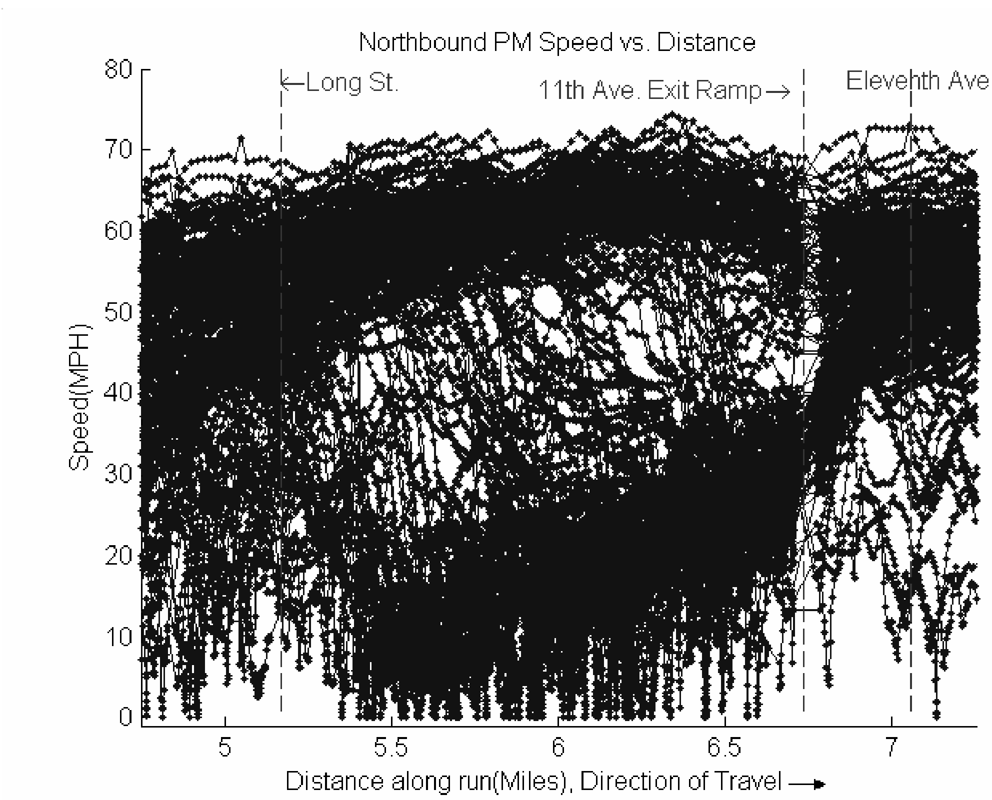

Figure 3: 173 Runs

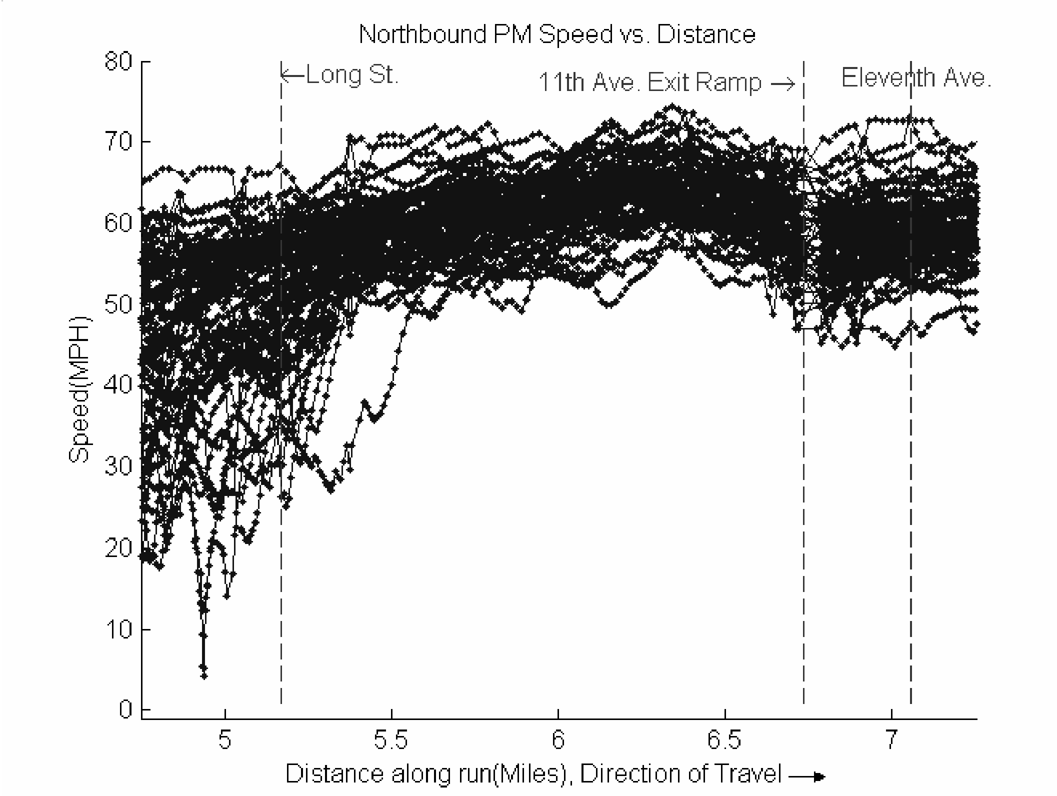

Figure 4: 71 Runs

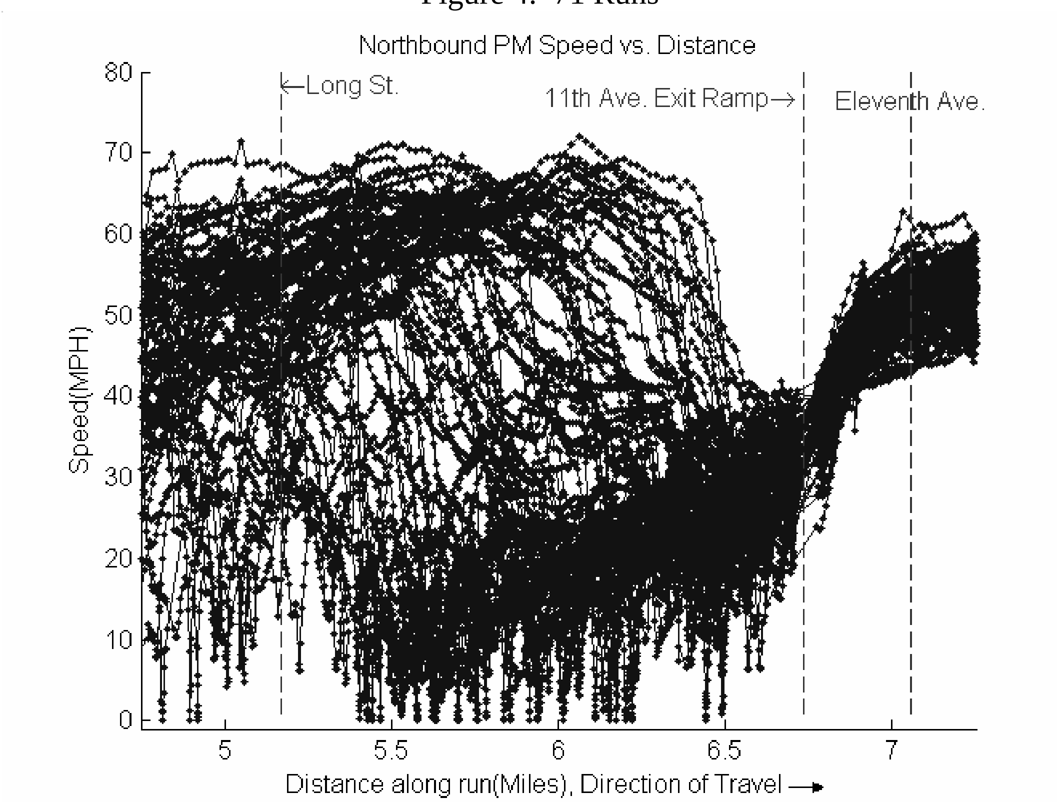

Figure 5: 91 Runs

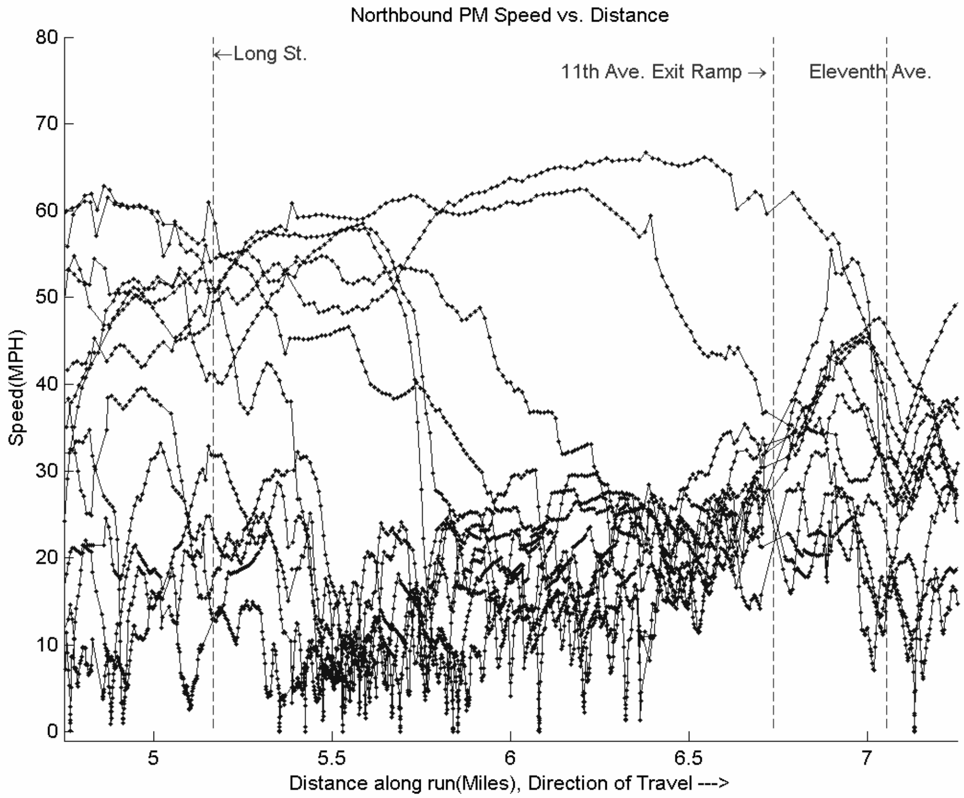

Figure 6: 12 Runs

Figure 7 illustrates probe vehicle runs passing a bottleneck around mile 5 on southbound I-71. As before, each of the curves on the plot is a different probe vehicle run. The probe vehicle travels south which corresponds to right to left on the figure. The traffic experiences a bottleneck similar to the northbound bottleneck: some of the runs do not experience congestion and remain free flowing, as shown in figure 8, some of the runs slow and enter the queue, and the queued runs ultimately pick up speed and return to free flow conditions after passing the bottleneck, shown in figure 9. The similarities between both the northbound and southbound bottlenecks are significant because they indicate that bottlenecks and congestion take on the same appearance in probe vehicle data independent of location. Probe vehicles can thus be a good method of looking for bottlenecks on a roadway.

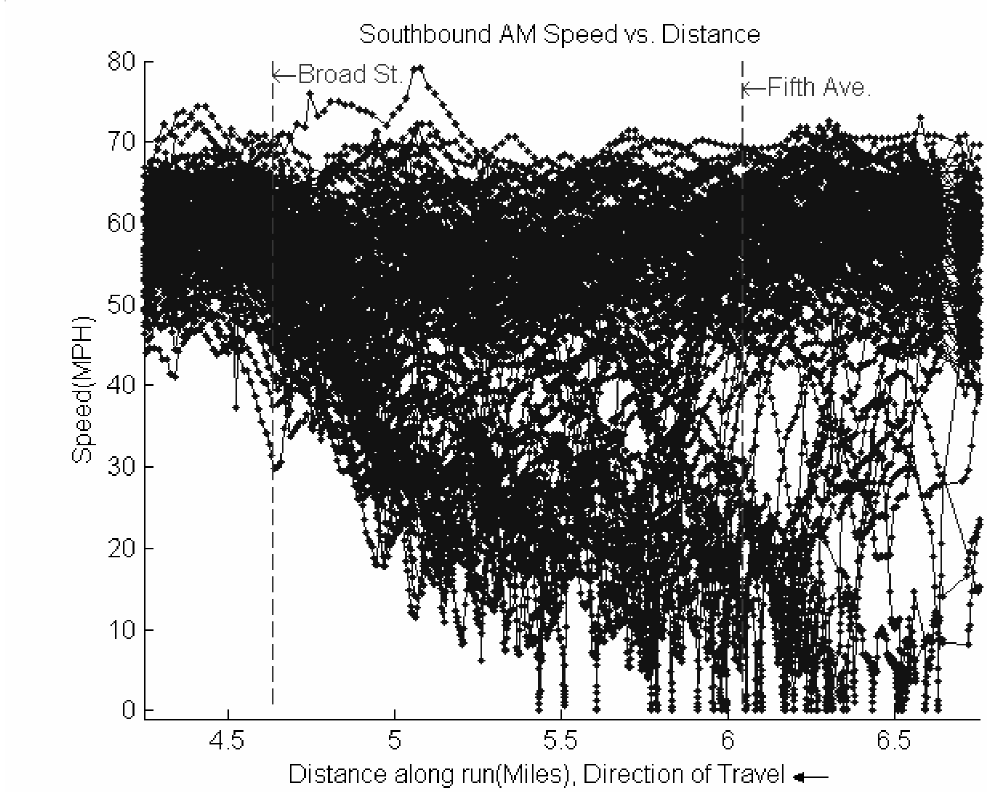

Figure 7: 122 Runs

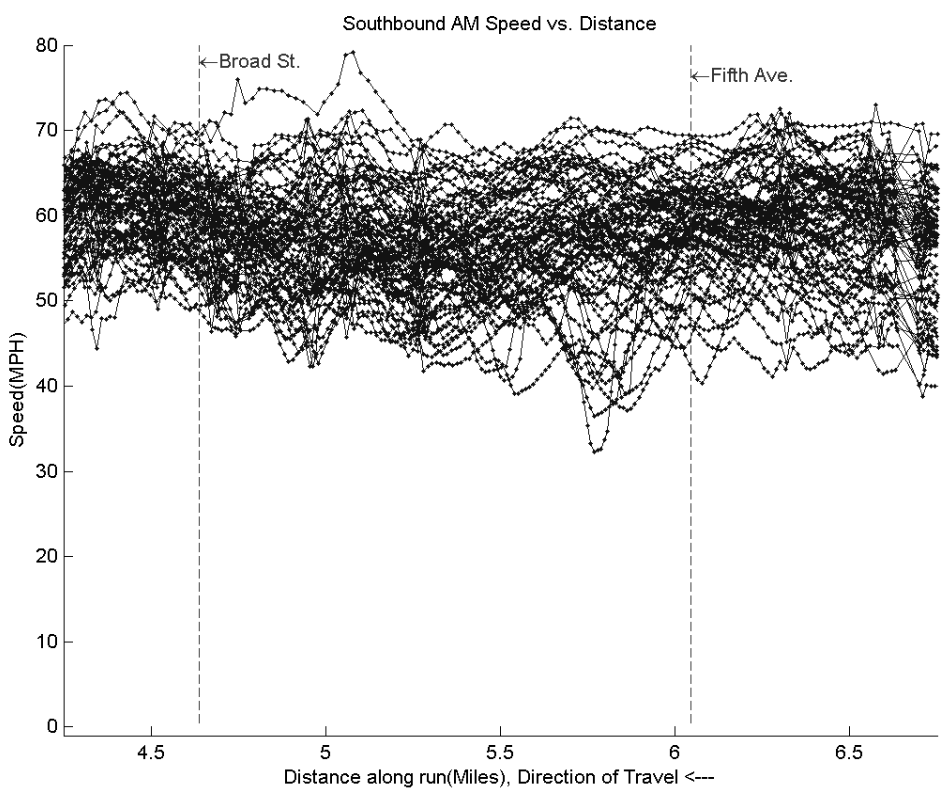

Figure 8: 76 Runs

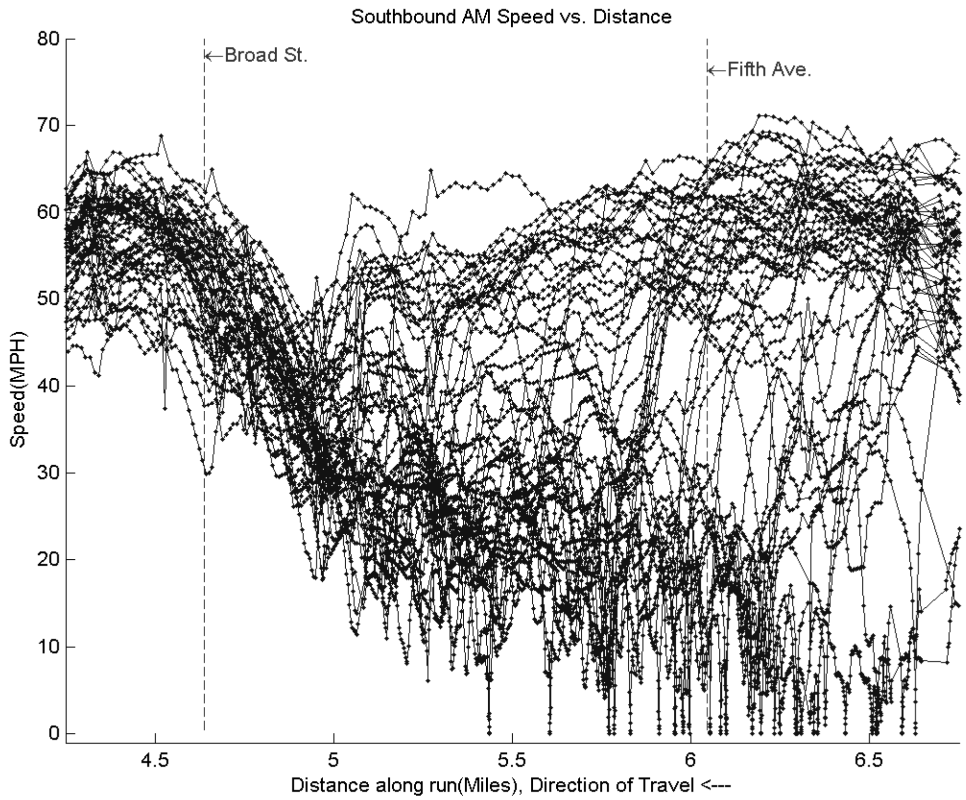

Figure 9: 46 Runs

Impacts of Sample Size Monte Carlo Simulation

The preceding figures show conditions changing day to day, e.g., a bottleneck may be active. A DGPS equipped probe vehicle can only provide an indication of how traffic is behaving and flowing in the vicinity of the probe. A single probe vehicle run may observe unusual conditions and if enough data are collected over a period of time, roadway trends such as bottleneck location, the typical amount of time it takes to travel a particular stretch of road (travel time), and average queue length can be established. While the number of runs used in this work is impractical for many studies, it would be useful to establish the quantity of data sufficient to determine roadway characteristics such as bottleneck locations and typical travel time. Fewer probe vehicle runs correspond to a lower cost to study a particular section of roadway.

In order to determine how many probe vehicle runs are required to study roadway characteristics, a Monte Carlo simulation was run. Northbound afternoon runs and southbound morning runs were the only runs considered because they are the most likely to experience recurring congestion and thus are not entirely free flowing. The northbound and southbound runs were broken down into travel links. Each link spanned I-71 from one interchange to the next. The links chosen for this study and link length are shown in table 1. Each link number is shown between a pair of roads; the roads are the boundaries of the links. The link locations are shown on the map in figure 10.

Table 1: Links of Study

Probe Vehicle's Direction of Travel ↑

Northbound Road/Link	Link Distance(Mi)
Hudson St.	
5	2.6
17th Ave.	
4	1.3
11th Ave.	
3	0.6
5th Ave.	
2	0.5
Long St	
1	1.2
70E/71N Join	

Probe Vehicle's Direction of Travel ↓

Southbound Road/Link	Link Distance (Mi)
Cooke Rd.	
1	0.8
North Broadway	
2	0.6
Weber Rd.	
3	0.6
Hudson St.	
4	1.3
17th St.	
5	1.1
5th St.	
6	1.5
Broad St.	
7	2.2
70W/71S Split	

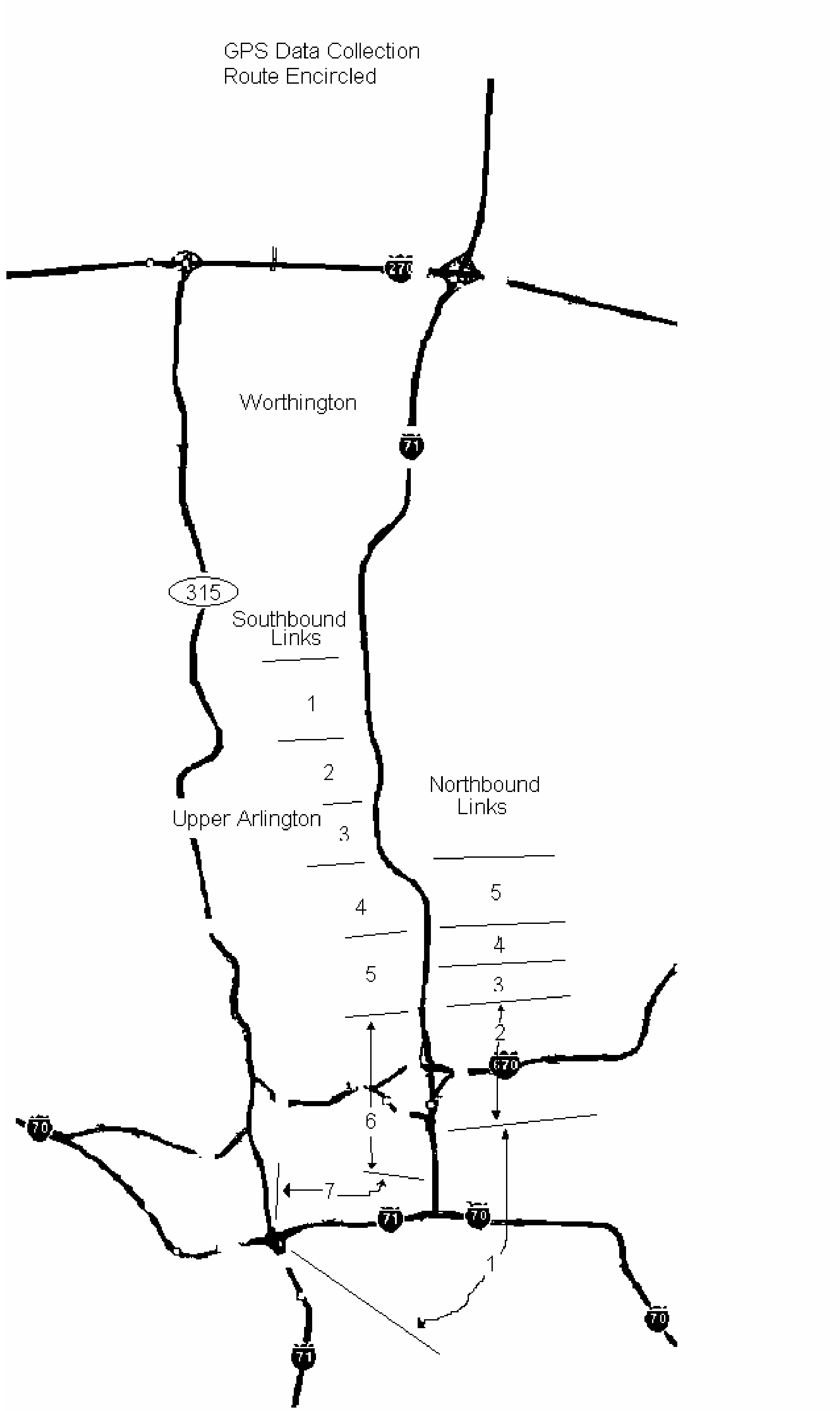

Figure 10: Illustration of Link Locations. Adapted from [2]

The algorithm for the Monte Carlo simulation is as follows:

1. Travel time for each link of the roadway on all probe vehicle runs was tabulated.
2. The median travel time of each link was found.
3. A set number of random probe vehicle runs were chosen out of the whole set of probe vehicle runs, either 5, 10, or 20 runs.
4. Travel time for each link of the roadway for the set of randomly chosen runs was tabulated.
5. The median travel time of each link was found for the randomly chosen set.
6. Steps 3-5 were repeated 1000 times for each of the 3 set sizes, and the results were compiled.
7. The median travel times of the 1000 randomly chosen sets at each number of runs were compared to the median travel time of the whole set.

Figure 11 shows the results of the simulation when only 5 probe vehicle runs are used to determine the link travel times of the roadway for the northbound afternoon runs. The solid line in the plot is the median travel time of each link from the full data set. The points above and below the solid line are the 5th percentile of the randomly chosen probe vehicle runs and the 95th percentile of the randomly chosen probe vehicle runs' travel time. It is interesting to note that with only 5 probe vehicle runs, travel time prediction is quite inaccurate. To quantify the range spanned by the simulated data, the decimal numbers in figure 6 are a representation of how far the 5th and 95th percentile travel times lie in relative terms from the median travel time and are defined as follows,

$$Range_{95}=(95^{th}\text{ Percentile TT}-5^{th}\text{ Percentile TT})/\text{Median TT},$$

where TT=Travel Time. The links which show high variation in $Range_{95}$ are the links which are regularly congested. The high variation implies that the variable length of queued traffic makes it difficult to accurately predict travel time with DGPS probe vehicle data. It should be noted that links 1 and 2 are longer in length than links 3-5, as shown in table 1, so links 1-2 show longer travel times. Links 2-3 show higher variability in Range95, which indicates frequent congestion and high variability in TT.

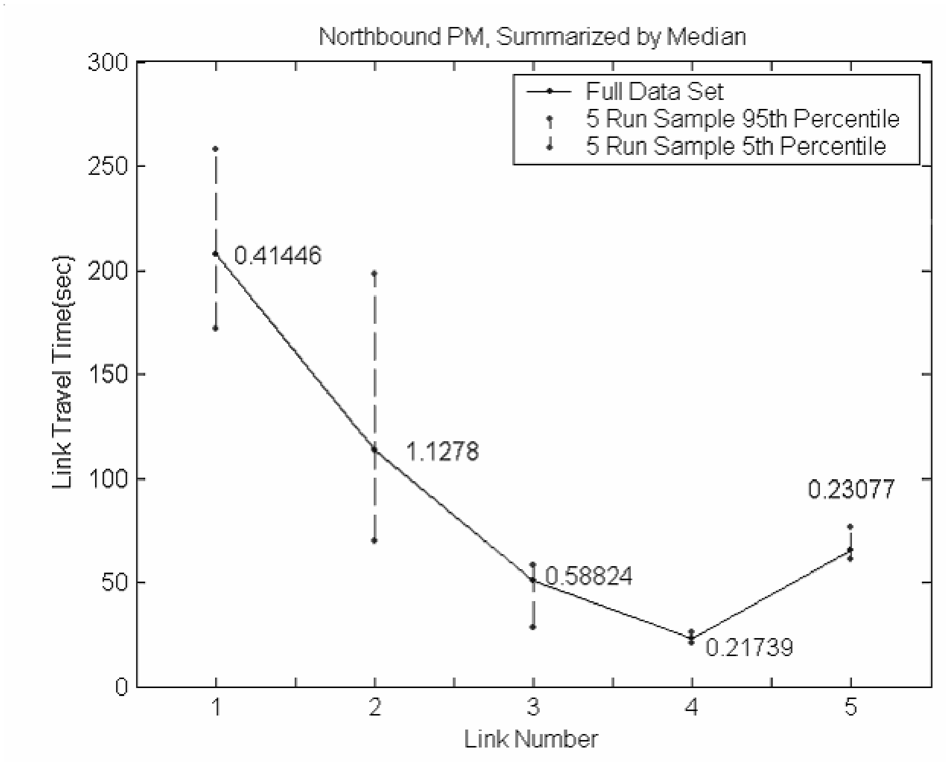

Figure 11

When the number of probe vehicle runs is increased to 20, as one might expect, the travel time quantification becomes more reliable, as shown in figure 12. Travel time for the links which tend to be uncongested (Links 1, 4, 5) can be more accurately predicted with 20 runs than can the more congested links (Links 2, 3). For completeness, figure 13 shows a plot when the number of probe vehicle runs is set at 10. As anticipated, $range_{95}$ for 10 samples falls between $range_{95}$ for 5 samples and 20 samples.

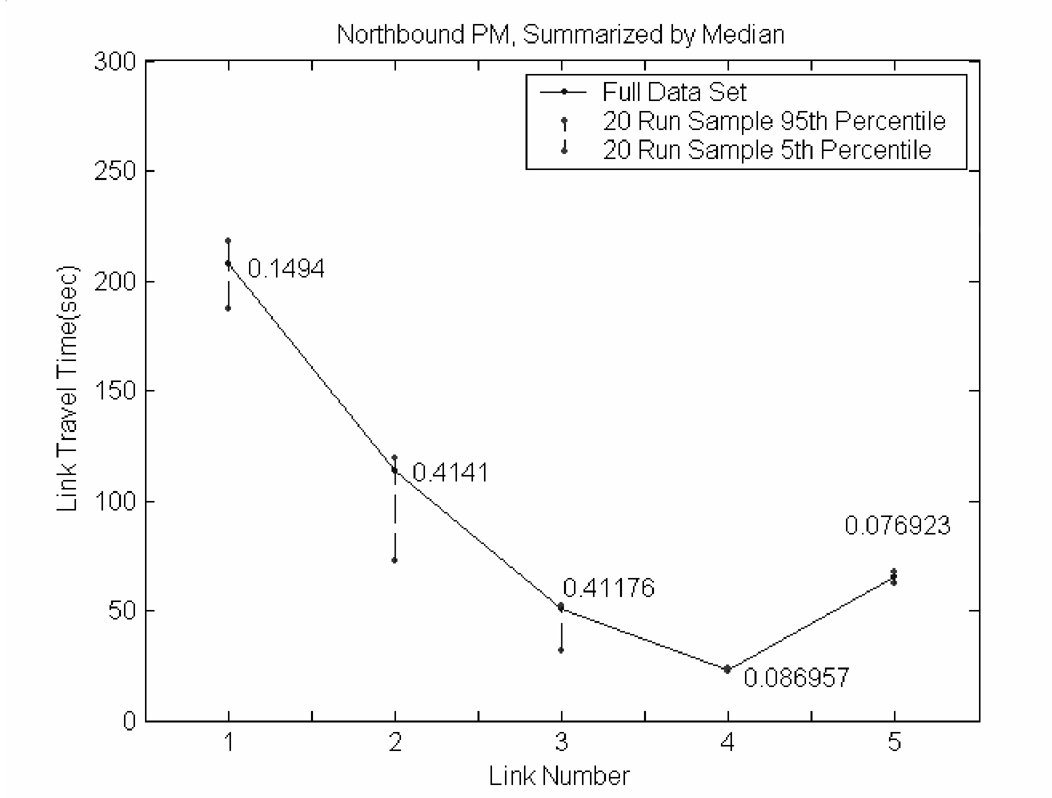

Figure 12

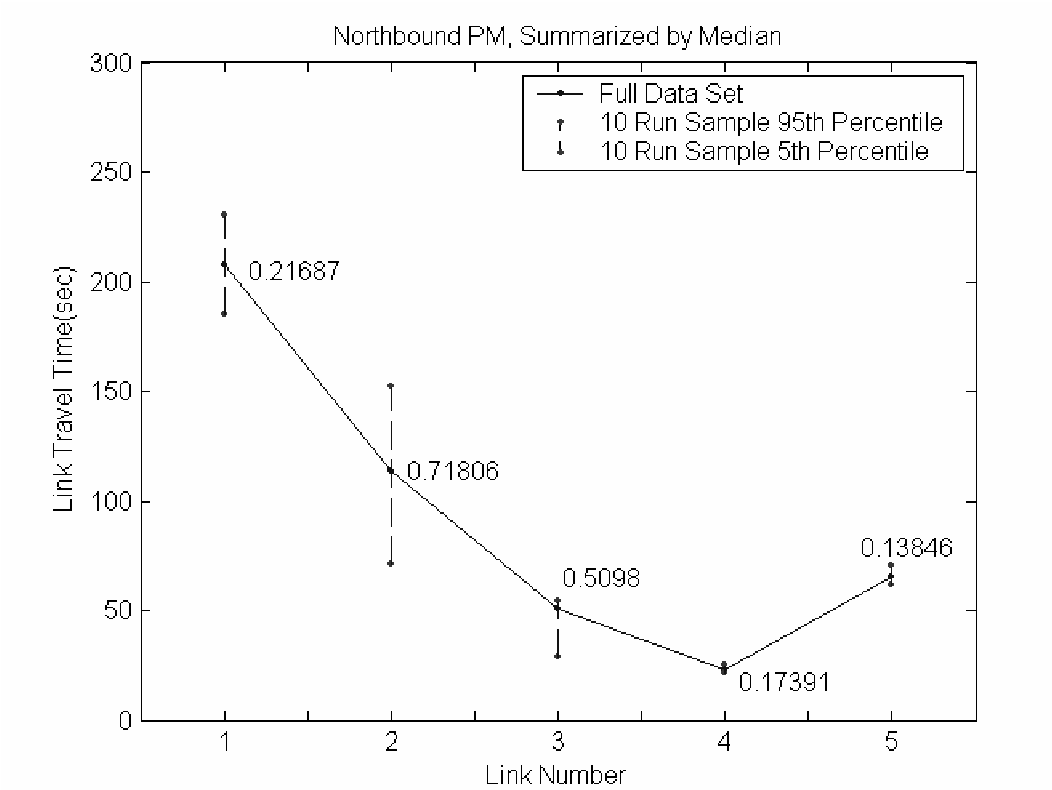

Figure 13

The southbound results are similar to the northbound results. With a sample size of 5 runs (figure 14), $range_{95}$ is large. When the sample size is increased to 10 runs (figure 15), $range_{95}$ is improved. Finally, when the sample size is increased to 20 runs (figure 16), $range_{95}$ is improved further.

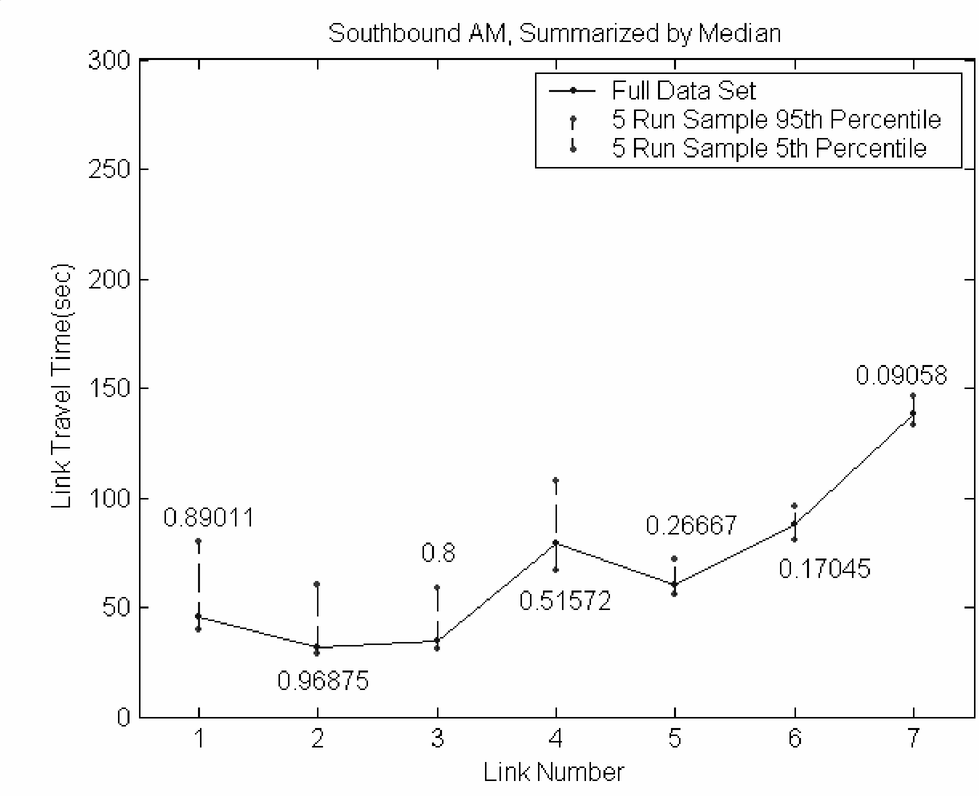

Figure 14

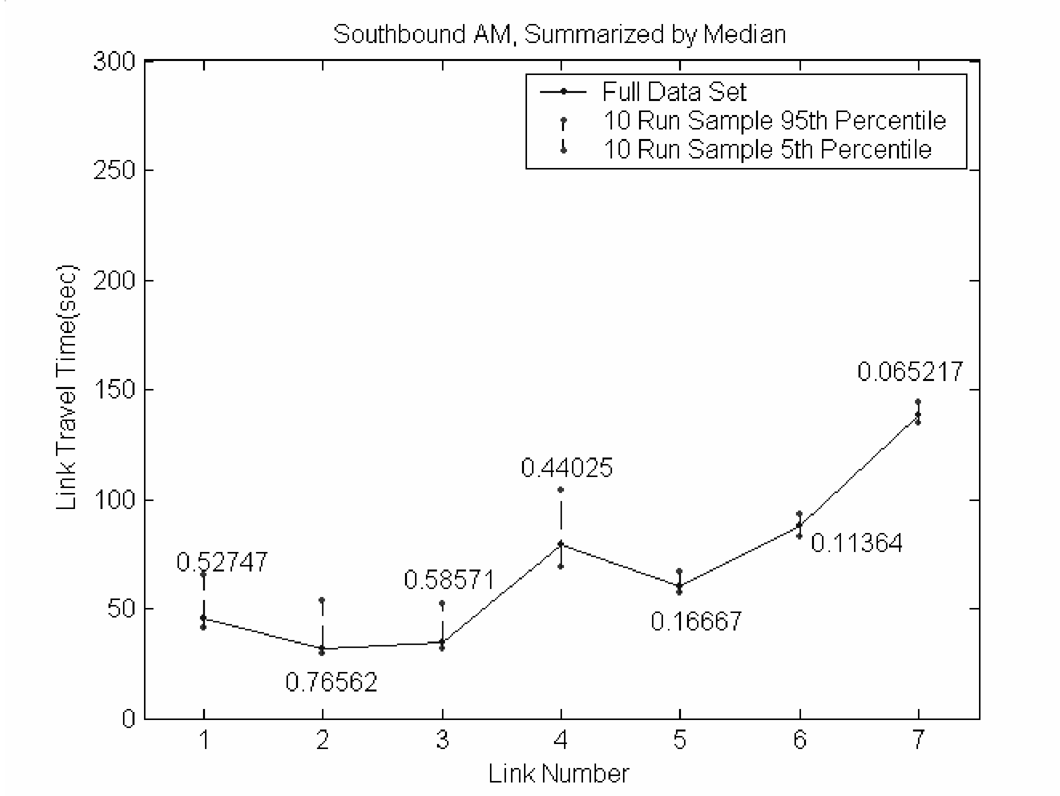

Figure 15

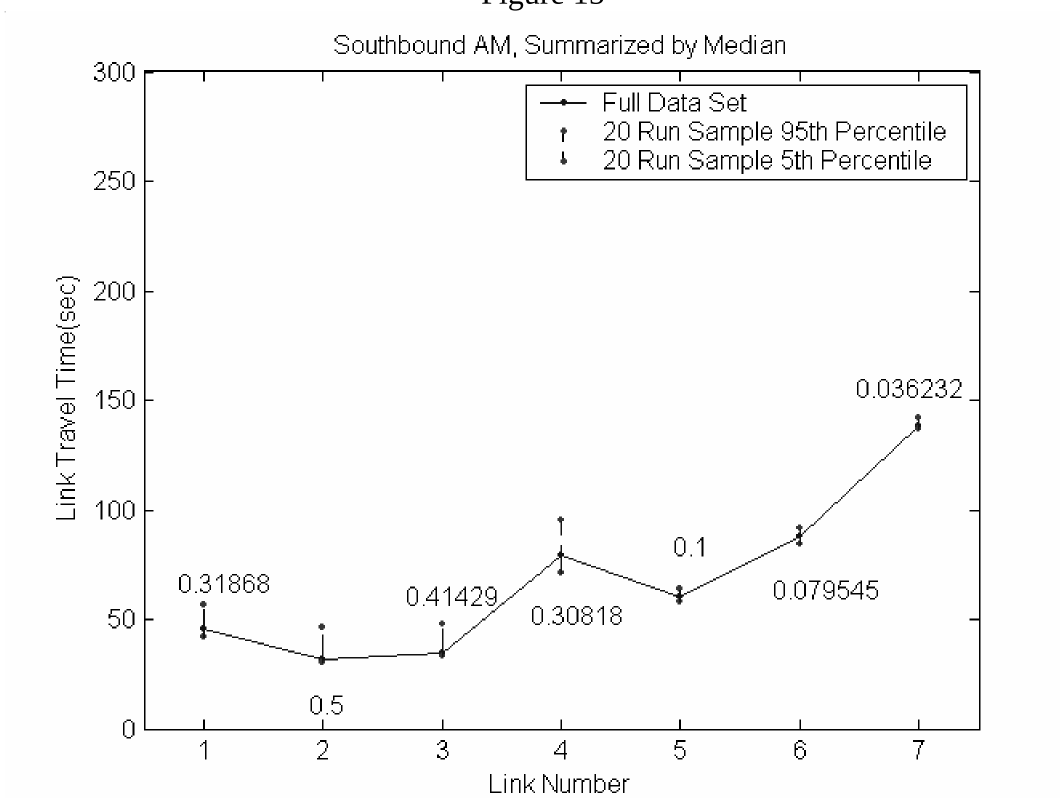

Figure 16

In order to better visualize how the number of runs per sample affects $Range_{95}$, the 5^{th} percentile and 95^{th} percentile TT vs. number of runs for each link are shown (northbound in figures 17-21, southbound in figures 22-28). Upon examination of the figures, one can see that increasing the number of runs past 20 provides only a small improvement in $Range_{95}$, if there is any improvement at all. The one exception where increasing the number of runs past 20 improves $Range_{95}$ is at northbound link 3 (figure 19), where increasing to 25 runs provides a somewhat better 5^{th} percentile value. Also note that northbound link 3 is the link most likely to be congested.

With only 3-5 probe vehicle runs during peak roadway usage hours, areas which experience frequent congestion can be located. In order to quantify the typical and maximum queue length and corresponding travel times with a high level of certainty in frequently congested areas, many probe vehicle runs (20+) may be needed. Once a probe vehicle run identifies an area of roadway that experiences congestion, other areas of the road that do not experience congestion could receive less attention so that the congested area can be studied in depth. In order to get as much information about a typically congested section of road with as few probe vehicle miles as possible, the probe vehicle could enter the roadway only a few miles upstream of the congested section and exit the roadway as soon as the congestion is passed. The probe vehicle would then turn around and return to keep driving through the congested traffic. In this way, the length of congested traffic can be studied, as well as the velocity of the congested traffic. By taking many passes through the congested section of road, the queue length can be quantified as it grows and shrinks through the roadway's peak usage hours.

Conclusion

Data from a DGPS equipped probe vehicle can be quite useful. The data can pinpoint the location of bottlenecks, which can then be correlated to roadway geometries and traffic flows, in an effort to correct the problems or prevent them from occurring at other locations. Also, the amount of congestion caused by a bottleneck can be quantified in terms of queue length without the expense of deploying more costly sensors, e.g. loop detectors. The low costs of GPS data collection could allow for cost benefit analysis of potential road improvements.

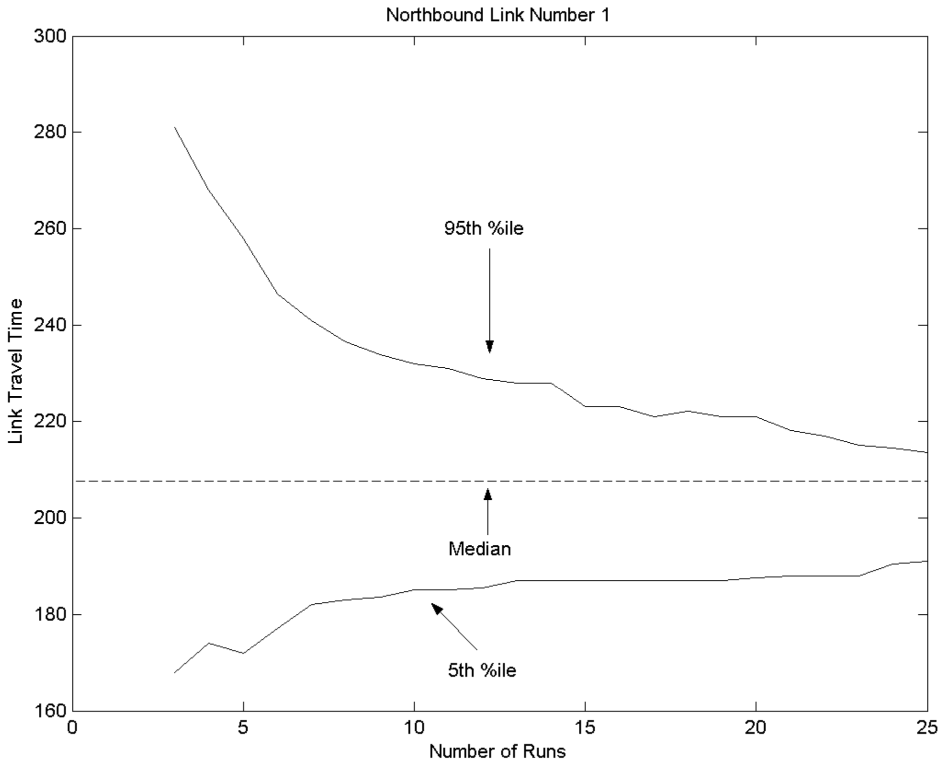

Figure 17

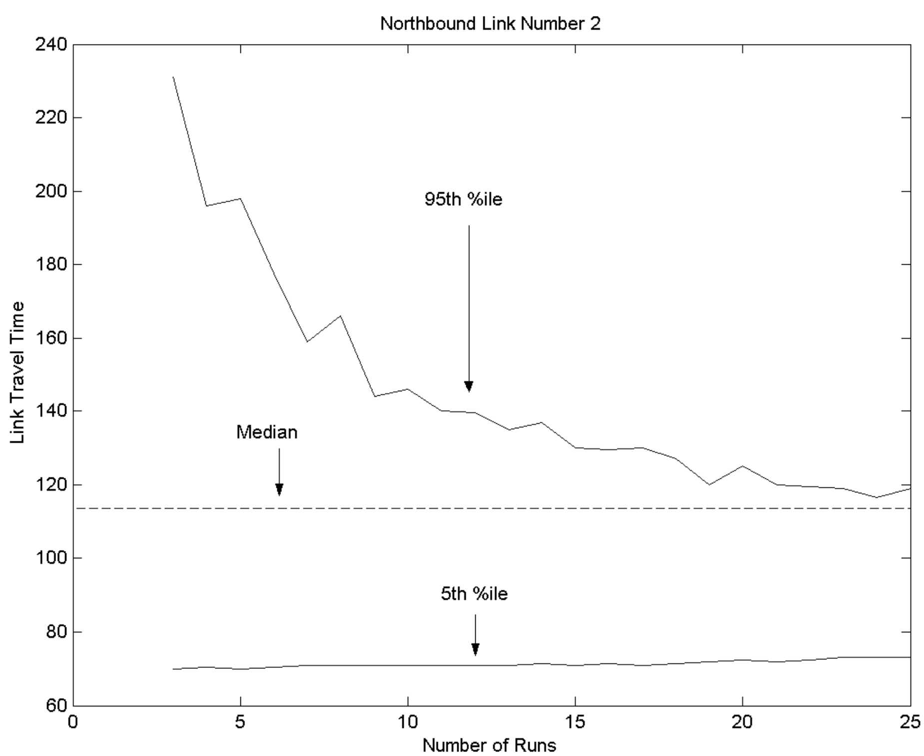

Figure 18

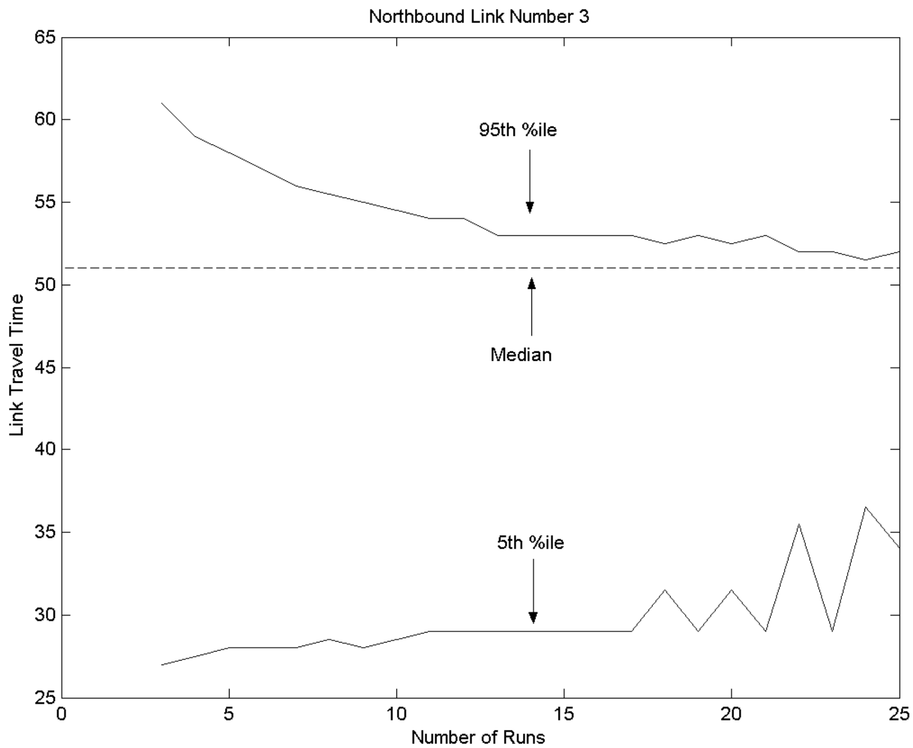

Figure 19

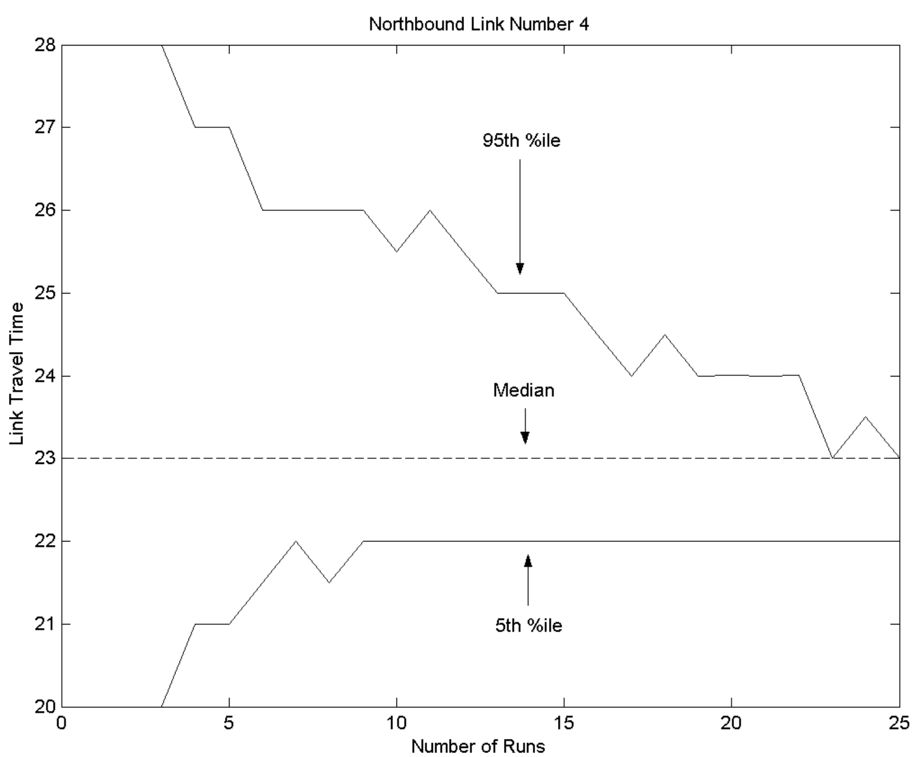

Figure 20

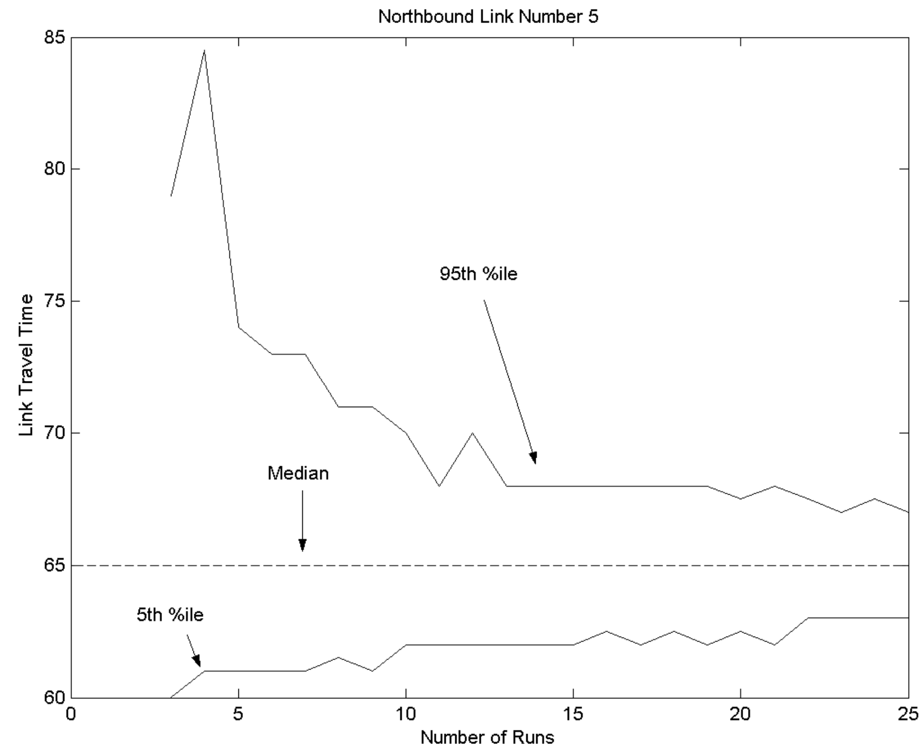

Figure 21

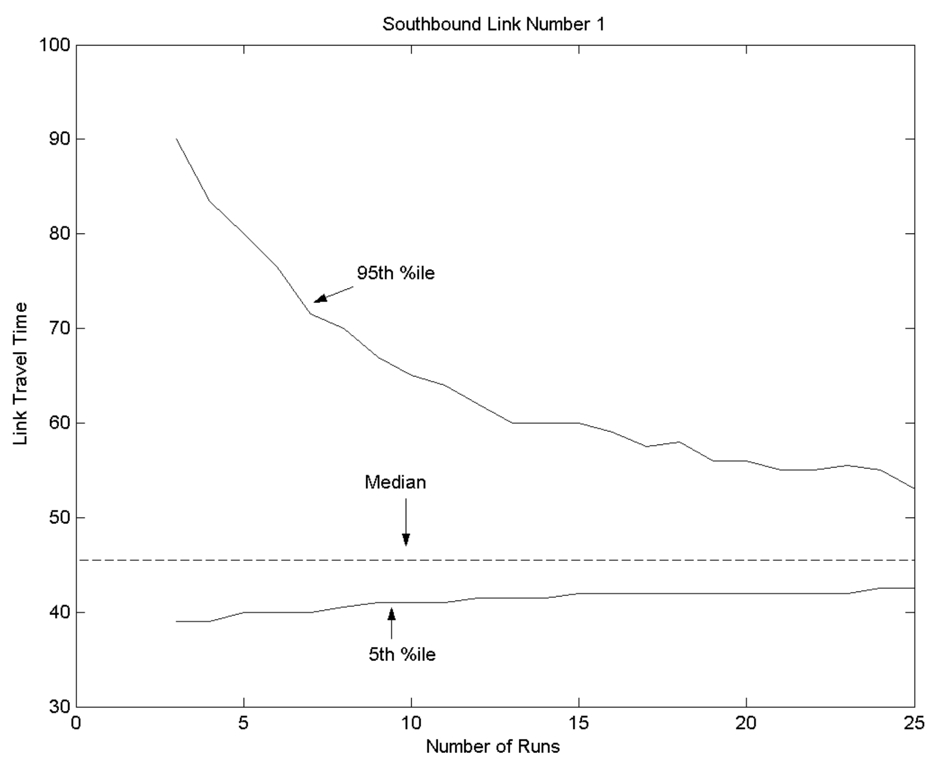

Figure 22

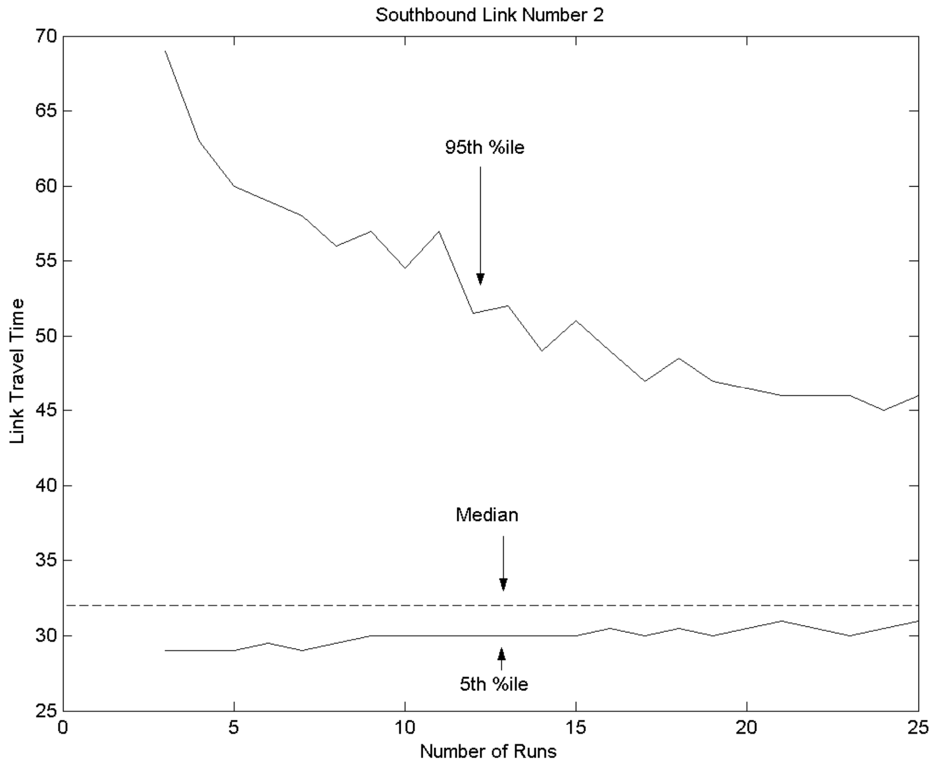

Figure 23

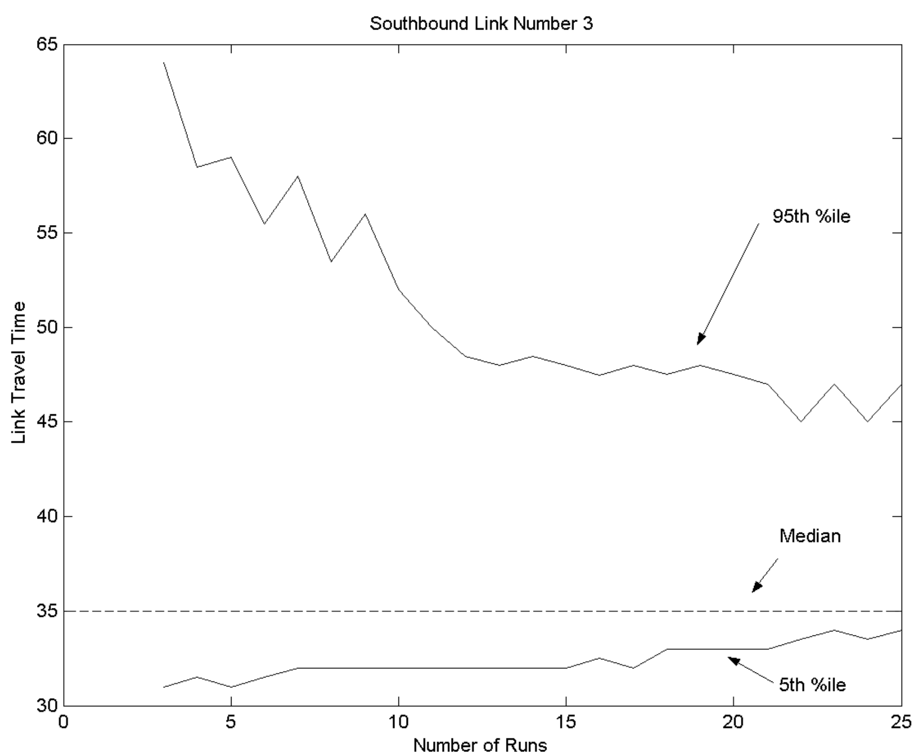

Figure 24

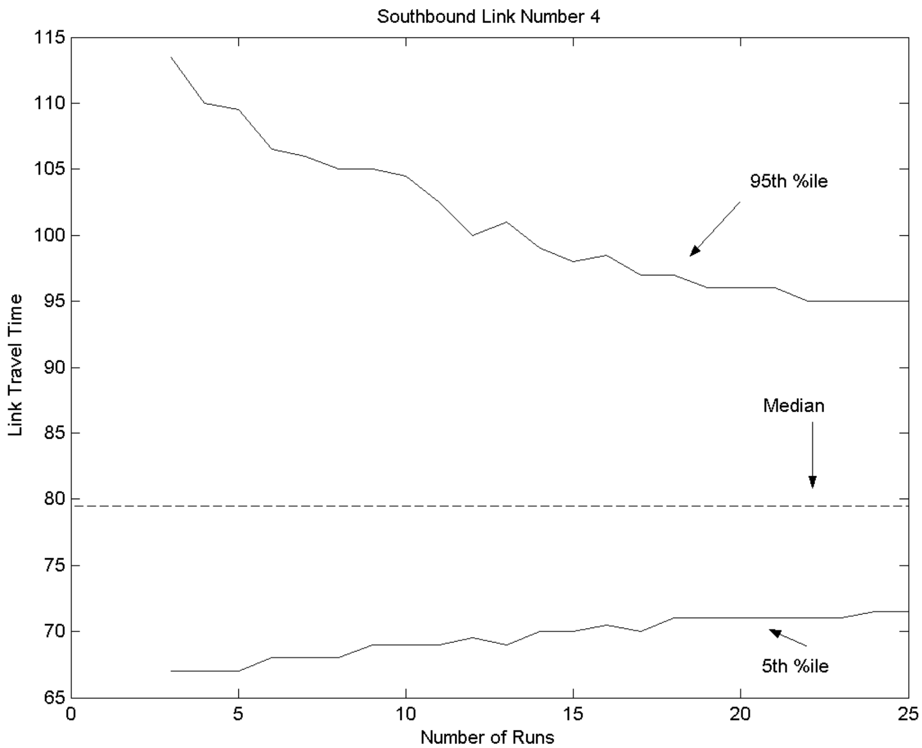

Figure 25

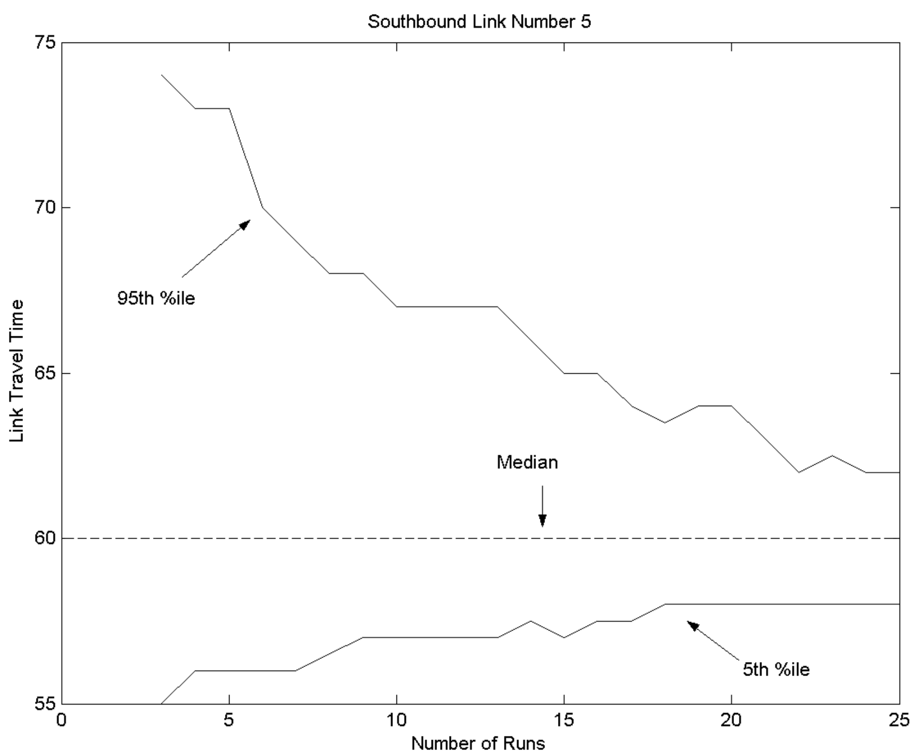

Figure 26

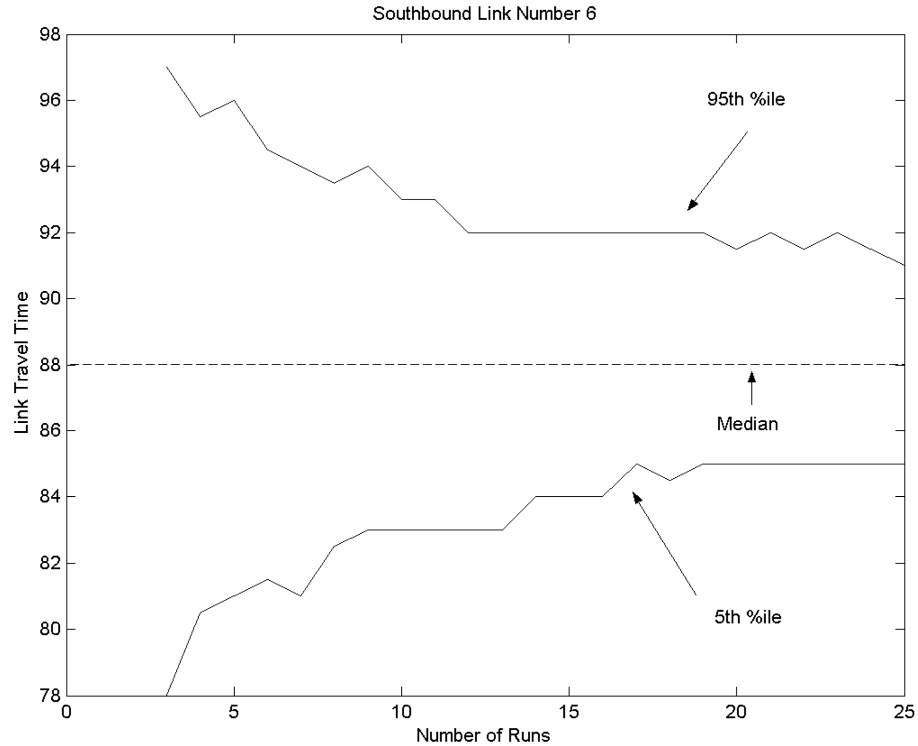

Figure 27

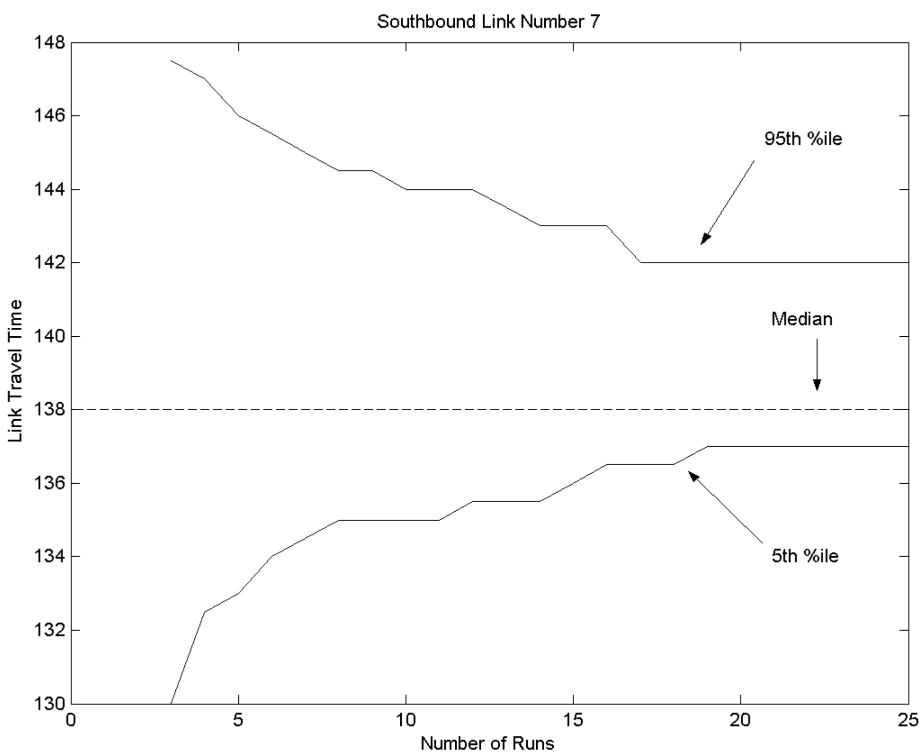

Figure 28

References

[1] Trimble Navigation Limited, “Differential GPS,” 2005; http://www.trimble.com/gps/dgps.html

[2] Mapquest.com, Inc., “Mapquest: Maps, Directions, and More,” 2005; http://www.mapquest.com

[3] S. Turner, W. Eisele, R. Benz, and D. Holdener, “Travel Time Data Collection Handbook,” FHWA-PL-98-035, Mar. 1998; http://www.fhwa.dot.gov/ohim/tvtw/natmec/00020.pdf

[4] J. Rice, E. Zwet, “A Simple and Effective Method for Predicting Travel Times on Freeways,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 3, September 2004.

Acknowledgements

I would like to thank Xin Wang for allowing me to use his code to snap the probe vehicle data to the reference run.